

Each-We Dilemmas and Effective Altruism

MATTHEW CLARK

University of St Andrews

THERON PUMMER

University of St Andrews

ABSTRACT

In his interesting and provocative article ‘Being Good in a World of Need’, Larry Temkin argues for the possibility of a type of Each-We Dilemma in which, if we each produce the most good we can individually, we produce a worse outcome collectively.¹ Such situations would ostensibly be troubling from the standpoint of Effective Altruism, the project of finding out how to do the most good and doing it, subject to not violating side-constraints (MacAskill, forthcoming, p. 5). We here show that Temkin’s argument is more controversial than it may appear initially regarding both impartiality and goodness. This is because it is both inconsistent with (i) a plausible conception of impartiality (Anonymity) and inconsistent with (ii) the standard view of goodness (the Internal Aspects View). Moreover, because (i) and (ii) are entailed by the sense of ‘impartial goodness’ that Effective Altruism tentatively adopts, Temkin’s argument is less relevant to Effective Altruism than he suggests.

1. FROM DISPERSE ADDITIONAL BURDENS TO EACH-WE DILEMMAS

Consider the following principle, which Temkin claims most people find ‘deeply compelling’ (Temkin, 2019, p. 13).

1. With Temkin, we are here concerned with the *impartial* goodness of outcomes (see Temkin, 2019, p. 13).

Disperse Additional Burdens: *In general, if additional burdens are dispersed among different people, it is better for a given total burden to be dispersed among a vastly larger number of people, so that the additional burden any single person has to bear within her life is ‘relatively small’, than for a smaller total burden to fall on just a few, such that their additional burden is substantial* (Temkin, 2019, p. 12; Temkin 2012, chapter 3).

Crucially, this principle is intended to make a claim about when one outcome is better, impartially speaking, than another.² Next consider Temkin’s case.

Reservoir: *Uhuru is walking by a reservoir where a child is drowning. If she pauses to remove her watch before diving in, the child will suffer severe brain damage. If she doesn’t remove her watch, its battery will leach toxic chemicals into the reservoir, increasing its pollution level by a very small amount [slightly burdening each of the 1,000,000 people who drink water from it]* (Temkin, 2019, p. 13).

Suppose that Uhuru is choosing between the child receiving an additional burden of size 10,000 ('substantial') and 1,000,000 people each receiving an additional burden of size 1 ('relatively small'). Independently of whether Uhuru *should* dive in immediately, Disperse Additional Burdens entails that she produces the *impartially best outcome* by doing so.

But now suppose Reservoir is iterated. Suppose there are 30,000 other people, each at the same reservoir as Uhuru, in a situation exactly similar to hers. That is, in front of each of these 30,000 people is a drowning child, and each could either dive in immediately thereby saving the child in front of them but increasing the reservoir's pollution level by a very small amount, or instead remove their watch before diving in, thereby avoiding polluting while saving the child but only after the child has suffered severe brain damage. According to Disperse Additional Burdens, while by diving in immediately *each* of these 30,000 people would produce the best outcome they can individually, they *together* produce an outcome which is worse than had they all removed their watches before diving in. For they together would be bringing about an additional burden of size 30,000 for each of the 1,000,000 people drinking

2. Note that Disperse Additional Burdens might instead be formulated as a deontic principle, concerning the choice-worthiness of actions rather than the betterness of outcomes. Ross has similarly argued, in response to Temkin, that Person-Affecting Views should be understood as deontic rather than axiological. For discussion, see: Ross, 2015; Temkin, 2015.

the reservoir water rather than bringing about an additional burden of size 10,000 for each of the 30,000 drowning children, and we can suppose that according to Disperse Additional Burdens the latter is better than the former.

If Disperse Additional Burdens is true, then there are situations in which, though each of us produces the impartially best outcome we can *individually*, we together produce an impartially worse outcome *collectively* (Temkin, 2019, p. 12). This is one sort of *Each-We Dilemma*.³

2. IS DISPERSE ADDITIONAL BURDENS PLAUSIBLE?

We contend that Temkin's reliance on Disperse Additional Burdens as a principle of impartial goodness both makes his argument more controversial than it may appear initially, and less relevant to Effective Altruism than he suggests. First, the initial plausibility of Disperse Additional Burdens is, at least in many cases, best explained by other less controversial principles. Moreover, Disperse Additional Burdens is both inconsistent with (i) a plausible conception of impartiality (Anonymity) and inconsistent with (ii) the standard view of goodness (the Internal Aspects View). As (i) and (ii) are entailed by the sense of 'impartial goodness' that Effective Altruism tentatively adopts, Temkin's argument is less relevant to Effective Altruism than he suggests.

There are several considerations, besides Disperse Additional Burdens, that may be able to account for our intuitions about Reservoir, and its iterated version. For example, our intuitions may be influenced by the view that as Uhuru is in an emergency rescue situation involving a face-to-face encounter with a child, she ought to dive straight in independently of whether she produces the impartially best outcome.⁴ Moreover, given realistic assumptions about the low probability of harm and costs of deliberation, the best strategy is to dive in unthinkingly. We are also subject to a mistaken general tendency to ignore small effects on large numbers of people (Parfit, 1984, ch.3).

In addition to difficulties finding positive support for Disperse Additional Burdens, we note that it is inconsistent with a plausible conception of impartiality,

3. Temkin borrows this term from Parfit, 1984, chapter 4 (note that Temkin uses 'Each-We Dilemma' in an arguably broader sense than Parfit).

4. For discussions of the potential moral relevance of distance and salience, see: Unger, 1996; Kamm, 2007; Woollard, 2015; Temkin, *The 2017 Uehiro Lectures*; Chappell, forthcoming; Mogensen, forthcoming.

according to which *who in particular* has a given well-being level makes no difference to which of two outcomes is better (that is, impartially better). On this conception, the outcome in which Ann is at level 1, Beth is at level 10, and Cathy is at level 100 is as good as any outcome in which *someone* is at level 1, *someone* is at level 10, and *someone* is at level 100, regardless of who in particular these people are. Call this conception of impartiality *Anonymity*.⁵

To see that *Disperse Additional Burdens* is inconsistent with *Anonymity*, compare the following two outcomes, A and B, each containing the same ten people, P₁ through P₁₀:⁶

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀
A	1	2	3	4	5	6	7	8	9	10
B	3	4	5	6	7	8	9	10	11	2

(Numbers in each row represent well-being levels.)

If *Anonymity* is correct, then the conclusion that B is better than A cannot plausibly be resisted. The worst-off person in B (P₁₀) is at a higher level than the worst-off person in A (P₁), the second worst-off person in B (P₁) is at a higher level than the second worst-off person in A (P₂), and so on. However, P₁₀ is much worse off in B than in A (losing 8 units of well-being) while each other person loses only 2 units of well-being if A comes about rather than B. Thus, assuming that a loss of 8 units is a ‘substantial’ burden whereas a loss of 2 units is ‘relatively small’, *Disperse Additional Burdens* entails that A is better than B. So, *Disperse Additional Burdens* is inconsistent with *Anonymity*.

The foregoing highlights one way in which *Disperse Additional Burdens* is more controversial than it may have appeared initially. Moreover, Effective Altruism is at least tentatively about promoting well-being *impartially*, that is, *counting everyone’s well-being equally*, in a way that respects *Anonymity* (MacAskill, forthcoming, p. 5).

5. Since *Anonymity* is about axiology only, it is compatible with ‘person-tracking’ non-consequentialist principles. For relevant discussion, see: Parfit, 2003; Broome, 2004, p. 135; Otsuka, 2018; and Brown, unpublished.

6. Inspired by Parfit, 2003, footnote 16 ‘musical chairs’; also see Temkin, 2012, pp. 440–445 ‘progressive disease’.

Disperse Additional Burdens is also inconsistent with the standard view of goodness, as given by the:

Internal Aspects View: For any outcome O, O has a unique degree of goodness, determined solely by O's internal features; and one outcome is better than another if and only if it has a higher degree of goodness (Temkin, 2012, p. 370).

In the example above, A's degree of goodness, as determined solely by its internal features, is lower than B's degree of goodness, as determined solely by its internal features. Therefore, according to the Internal Aspects View, B is better than A. This is inconsistent with Disperse Additional Burdens.

Over many years, Temkin has argued against the Internal Aspects View, and instead defended the *Essentially Comparative View*, according to which, for at least some outcomes, the degree of goodness of an outcome is relativized to comparisons, i.e. there is no fact about how good an outcome is *on its own*, but only relative to what other outcomes it is compared with (Temkin, 2012, p. 371). But we believe that the Internal Aspects View is the standard view for good reason. It is the sort of view that nearly everyone finds very plausible, upon grasping it (Huemer, 2013).

It should accordingly be accepted, absent sufficiently strong countervailing considerations. For reasons we cannot explain here, we doubt anyone has presented such considerations.⁷

In his article Temkin suggests a further line of support for Disperse Additional Burdens. He presents his *Lollipops for Life* case, suggesting most people judge, contrary to total utilitarianism,⁸ that an outcome containing a miserable innocent person alongside countlessly many well-off people is worse than an outcome containing a flourishing innocent person alongside the very same countlessly many well-off people, where they each get one fewer lollipop lick. However, standard explanations of such judgements—that no number of lollipop licks is as good as a flourishing life,⁹ that the well-being of the least well off is of special importance for the goodness of out-

7. For responses to Temkin's 'spectrum arguments' against the Internal Aspects View, see: Pummer, 2018; and Nebel, 2018.

8. Total utilitarianism holds that one outcome is better than another if and only if it contains more well-being in total.

9. Consider, for example, 'value superiority' views in the tradition of Mill's claims about higher and lower pleasures. For relevant literature, see: Arrhenius and Rabinowicz, 2015; Parfit, 2016; and Clark, unpublished.

comes¹⁰—are consistent with the Internal Aspects View of goodness. Consequently, we find it somewhat misleading when Temkin writes, ‘[f]or certain comparisons, at least, [most people] reject total utilitarianism’s simple additive-aggregationist approach in favour of the anti-additive-aggregationist approach of principles like the Disperse Additional Burdens View’ (Temkin, 2019, p. 13). Even if Lollipops for Life yields strong reason to reject total utilitarianism’s simple additive-aggregationist approach, it does not give us compelling reason to accept Disperse Additional Burdens over the Internal Aspects View of goodness.

Finally, Effective Altruists are concerned with quantities like ‘number of lives saved’, ‘number of children dewormed’, and so on, as *amounts of good*, which do not depend on what alternatives are available. These amounts of good feature centrally in cost-effectiveness comparisons across different sets of alternatives, and combine with subjective probabilities in expected utility calculations. The Effective Altruist project, so construed, presupposes the Internal Aspects View. At the same time, it does *not* presuppose total utilitarianism, or what Temkin calls an ‘additive-aggregationist approach’.

3. EFFECTIVE ALTRUISM AND COORDINATION

Nothing we have argued implies that there are no practical issues of coordination, or of risks of systemic negative side-effects, that are of extreme relevance to Effective Altruism.¹¹ On the contrary, these and related issues are rightly very much on the minds of Effective Altruists. Especially in recent years, Effective Altruists have been developing solutions to coordination problems, and building mechanisms of effectively ‘doing good together’ (Dietz, 2018; Collins, forthcoming).

Coordination problems can indeed give rise to Each-We Dilemmas where, if we each produce the best outcome we can individually, we produce a worse outcome collectively. To see this, consider the following familiar example (Gibbard, 1965):

¹⁰. Consider the “well-known anti-additive-aggregationist principles of equality and maximin” (Temkin, 2012, p. 70).

¹¹. On risks of harm, see: Wenar, 2011; and Pummer, 2016.

		You
	Do nothing	Do X
I	Do nothing Second-best (0)	Bad (-10)
	Do X Bad (-10)	Best (10)

If we both do nothing, we each produce the best outcome we can *individually* (0), as doing X would produce a worse outcome (-10). But we together produce a worse outcome (0) *collectively* than had we both done X (10). This sort of Each-We Dilemma can be solved by improving coordination in the sense that if we each do the action that produces the best outcome *collectively*, then each of us also produce the best outcome we can *individually*.

Temkin's Each-We Dilemmas cannot be solved in this way. In these situations, the *only* outcomes that can be produced when we each produce the best outcome *individually* are those that are worse *collectively*. But these are precisely the sorts of situations that Regan, Parfit, and others working on coordination problems, have claimed are impossible.¹² Indeed, they are impossible if the Internal Aspects View is true. The Internal Aspects View entails that there is an outcome we could together produce that is *not worse collectively* than any other we could together produce (since, according to this View, each outcome has a unique degree of goodness).¹³ Now assume that we together produce an outcome O that is not worse *collectively* than any other we could together produce. If, by acting in some other way, one of us could produce outcome O+ which is better *individually*, then, according to the Internal Aspects View, outcome O+ is also better *collectively* and we could together produce it. But this contradicts the assumption that O is not worse *collectively* than any other outcome we could together produce.

The putatively troubling Each-We Dilemmas implied by Temkin's argument will not arise for projects that promote impartial goodness as construed by the Internal Aspects View. Effective Altruism is such a project. Consequently, Temkin's argument is less relevant to Effective Altruism than he suggests.

12. For discussion, see: Regan, 1980, pp. 54–55; Parfit 1984, especially section 21 and p. 91; Rabinowicz, 1989; and Temkin 2012 chapter 3, footnote 19.

13. Assuming there is a finite number of available outcomes.

In closing, we would like to emphasize that Temkin does not believe it is *wrong* to individually pursue the Effective Altruist project, as we construe it. In personal communication, dated 27 July 2018, he wrote:

[O]n my view, though doing that is likely to be both permissible and laudable, relative to most other things one might do, I don't believe that doing so will also be the MOST laudable thing to do. Nor do I believe that a truly good person will, or should so far as possible, always or even generally be motivated by that project. But surely there will sometimes, and perhaps even often, be times when acting in such a way WOULD be precisely the thing that a good person would, and should, do.

We agree that pursuing the Effective Altruist project, as we construe it, is permissible and laudable. We go further in advocating for its wider adoption, but it is, in any case, important to recognize the scope of the challenge Temkin raises.

Acknowledgements

For helpful comments, we are grateful to Roger Crisp, Tom Douglas, Ben Sachs, and Larry Temkin.

REFERENCES

- Arrhenius, Gustaf & Rabinowicz, Wlodek (2015). Value Superiority. In Iwao Hirose & Jonas Olson (eds.), *The Oxford Handbook of Value Theory*. New York: Oxford University Press.
- Broome, John (2004). *Weighing Lives*. Oxford: Oxford University Press.
- Brown, Campbell (unpublished). Anonymity and Moral Equality.
- Chappell, Richard (forthcoming). Overriding Virtue. In Hilary Greaves & Theron Pummer (eds.), *Effective Altruism: Philosophical Issues*. Oxford: Oxford University Press.
- Clark, Matthew (unpublished). Continuous Superiority.
- Collins, Stephanie (forthcoming). Beyond Individualism. In Hilary Greaves & Theron Pummer (eds.), *Effective Altruism: Philosophical Issues*. Oxford: Oxford University Press.
- Dietz, Alexander (2019). Effective Altruism and Collective Obligations. *Utilitas* 31 (1):106-115.
- Gibbard, Allan. (1965). Rule-utilitarianism: Merely an illusory alternative? *Australasian Journal of Philosophy* 43 (2):211-220.

- Huemer, Michael (2013). Transitivity, Comparative Value, and the Methods of Ethics. *Ethics* 123 (2):318-345.
- Kamm, Frances (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.
- MacAskill, William (forthcoming). The Definition of Effective Altruism. In Hilary Greaves & Theron Pummer (eds.), *Effective Altruism: Philosophical Issues*. Oxford: Oxford University Press.
- Mogensen, Andreas (forthcoming). The Callousness Objection. In Hilary Greaves & Theron Pummer (eds.), *Effective Altruism: Philosophical Issues*. Oxford: Oxford University Press.
- Nebel, Jacob. (2018). The Good, the Bad, and the Transitivity of Better Than. *Noûs* 52 (4):874-899.
- Otsuka, Michael (2018). How it makes a moral difference that one is worse off than one could have been. *Politics, Philosophy and Economics* 17 (2):192-215.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- (2003). Justifiability to each person. *Ratio* 16 (4):368-390.
- (2016). Can We Avoid the Repugnant Conclusion? *Theoria* 82 (2):110-127.
- Pummer, Theron (2016). Risky Giving. *The Philosophers' Magazine* 73 (2):62-70.
- (2018). Spectrum arguments and hypersensitivity. *Philosophical Studies* 175 (7):1729-1744.
- Rabinowicz, Wlodek (1989). Act-utilitarian prisoner's dilemmas. *Theoria* 55 (1):1-44.
- Regan, Donald. (1980). *Utilitarianism and Co-Operation*. Oxford: Oxford University Press.
- Ross, Jacob (2015). Rethinking the Person-Affecting Principle. *Journal of Moral Philosophy* 12 (4):428-461.
- Temkin, Larry (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press.
- (2015). Rethinking Rethinking the Good. *Journal of Moral Philosophy* 12 (4):479-538.
- (2017). *The 2017 Uehiro Lectures*, The Oxford Uehiro Centre for Practical Ethics.
- (2019). Being Good in a World of Need: Some Empirical Worries and an Uncomfortable Philosophical Possibility. *Journal of Practical Ethics* 7(1): 1-24.
- Unger, Peter (1996). *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.
- Wenar, Leif (2011). Poverty is No Pond: Challenges For the Affluent. In Patricia Illingworth, Thomas Pogge & Leif Wenar (eds.), *Giving Well: The Ethics of Philanthropy*. pp. 104-132. New York: Oxford University Press.
- Woollard, Fiona (2015). *Doing and Allowing Harm*. Oxford: Oxford University Press.